D4

OPIC
OFFICE DE LA PROPRIÉTÉ INTELLECTUELLE DU CANADA

CIPO
CANADIAN INTELLECTUAL PROPERTY OFFICE

(12) (19) (CA) **Demande-Application**

(21) (A1) **2,325,806**

(22) 2000/12/01
(43) 2001/06/03

(72) STEEG, EVAN W., CA
(72) MISENER, STEPHEN R., CA
(72) WILLIS, EDWARD S., CA
(72) ANDERSON, GARY E., CA
(72) DUNNE, RODERICK J., CA
(72) LEPAGE, MARC A., CA
(71) MOLECULAR MINING CORPORATION, CA

(51) Int.Cl.[7] C12Q 1/68, G01N 33/574
(30) 1999/12/03 (60/168,625) US

(54) **METHODES DE DIAGNOSTIQUE ET DE PRONOSTIC DE LEUCEMIES AIGUES**

(54) **METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF ACUTE LEUKEMIAS**

(57) The present invention relates to the diagnosis of the distinction between acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) and prognosis of AML. Disclosed is a means to diagnose the distinction between ALL and AML employing measurement of the abundance of the nucleic acid or protein products of small combinations (two, three or more) of particular human genes. The invention further describes the use of the measurement of the abundance of the nucleic acid or protein product of two human genes for prognostic indication in AML. The invention also relates to therapies targeted at these indicator genes, and the screening of drugs for cancer that target these indicator genes or their protein products.

DOCID: <CA__2325806A1_I_>

Methods for the Diagnosis and Prognosis of Acute Leukemias

Background of the Invention

Field of the Invention

The present invention relates to methods of classifying acute leukemias. More particularly, the invention relates to methods of distinguishing acute myeloid leukemia (AML) from acute lymphoblastic leukemia (ALL) by measuring the nucleic acid levels or gene product (protein) levels of small combinations (two, three or more) of particular human genes. The invention is also useful as a prognostic indicator in AML.

Related Art

A major challenge of cancer treatment has been to target specific therapies to pathogenically distinct tumor types, to maximize efficacy and minimize toxicity. Improvements in cancer classification have thus been central to advances in cancer treatment.

Cancer classification has been based primarily on morphological appearance of the tumor, but this has serious limitations. Tumors with similar histopathological appearance can follow significantly different clinical courses and show different responses to therapy. In a few cases, such clinical heterogeneity has been explained by dividing morphologically similar tumors into subtypes with distinct pathogeneses. Key examples include the subdivision of acute leukemias, non-Hodgkin's lymphomas, and of childhood "small round blue cell tumors" into neuroblastomas, rhabdomyosarcoma, Ewing's sarcoma, and other types. For many more tumors, however, important subclasses are likely to exist but have yet to be defined by molecular markers. For example, prostate cancers of identical grade can have widely variable clinical courses, from indolence over decades to explosive growth causing rapid patient death.

Cancer classification has been difficult in part because it has historically relied on specific biological insights, rather than systematic and unbiased approaches for recognizing tumor subtypes.

Acute leukemia is a disease of the leukocytes and their precursors. It is characterized by the appearance of immature, abnormal cells in the bone marrow and peripheral blood and frequently in the liver, spleen, lymph nodes, and other parenchymatous organs. The clinical picture is marked by the effects of anemia, which is usually severe (fatigue, malaise), an absence of functioning granulocytes (proneness to infection and inflammation), and thrombocytopenia (hemorrhagic diathesis). The spleen and liver usually are moderately enlarged, while enlarged lymph nodes are seen mainly in the pediatric lymphoblastic leukemias. Fever and a very high ESR complete the picture. Leukocyte counts vary greatly in the acute leukemias. About one-fourth to one-third of cases begin with a low white blood count (sub- or aleukemic leukemia), while about half show some degree of leukocytosis. Mature granulocytes may still be found in the peripheral blood in addition to abnormal forms. The coexistence of immature and mature cell forms is termed "hiatus leucaemicus." The leukocytopenic forms are the most difficult to differentiate from aplastic anemias, pancytopenias, and the myelodysplastic syndromes. Bone marrow aspiration is usually necessary to establish a diagnosis. Aspirated marrow is found to be permeated by abnormal cells (paramyeloblasts, paraleukoblasts, nonclassifiable cells (N.C.), leukemic cells, blasts, etc.) with little or no evidence of normal hematopoiesis.

The acute leukemias have traditionally been classified according to morphologic, cytochemical, and/or immunologic criteria. An overview of acute leukemia classification can be found in the "Atlas of Acute Leukemia" available on the world wide web at www.meds.com/leukemia/atlas/acute-leukemia.html.

As a brief historical review, the classification of acute leukemias began with the observation of variability in clinical outcome (Farber, S., *et al.*, *N. Engl. J. Med. 238*:787 (1948)) and subtle differences in nuclear morphology (Forkner, C.E., *Leukemia and Allied Disorders* , MacMillan, New York (1938); Frei, E., *et al., Blood 18*:431 (1961); Medical Research Council, *Br. Med. J. 1*:7 (1963)).

Enzyme-based histochemical analysis were introduced in the 1960s to demonstrate that some leukemias were periodic acid-Schiff positive, whereas others were myeloperoxidase positive (Quaglino, D., and Hayhoe, F.G.J., *J. Pathol 78*:521 (1959); Bennett, J.M., Dutcher, T.F., *Blood 33*:341 (1969); Graham, R.C., *et al., J. Histochem, Cytochem 13*:150 (1965)). This provided the first basis for classification of acute leukemias into those arising from lymphoid precursors (acute lymphoblastic leukemia, ALL) or from myeloid precursors (acute myeloid leukemia, AML). This classification was further solidified by the development in the 1970s of antibodies recognizing either lymphoid or myeloid cell surface molecules (Tsukimoto, I., *et al., N. Eng. J. Med. 294*:245 (1976); Schlossman, S.F., *et al., Proc. Natl. Acad. Sci. U.S.A. 73*:1288 (1976); Roper, M., *et al., Blood 61*:830 (1983); Sallan, B.S.E., *et al., Blood 55*:395 (1980); Pesando, J.M., *et al., Blood 54*:1240 (1979)). Most recently, particular subtypes of acute leukemia have been found to be associated with specific chromosomal translocations—for example, the t(12;21)(p13;q22) translocation occurs in 25% of patients with ALL, whereas the t(8;21)(q22;q22) occurs in 15% of patients with AML (Golub, T.R., *et al., Proc. Natl. Acad. Sci. U.S.A. 92*:4917 (1995); McLean, T.W., *et al., Blood 88*:4252 (1996); Shurtleff, S.A., *et al., Leukemia 9*:1985 (1995); Romana, S.P., *et al., Blood 86*:4263 (1995); Rowley, J.D., *Ann. Genet. 16*:109 (1973)).

Although the distinction between AML and ALL has been well-established, no single test is currently sufficient to establish the diagnosis. Rather, current clinical practice involves an experienced hematopathologist's interpretation of the tumor's morphology, histochemistry, immunophenotyping, and cytogenetic analysis, each performed in a separate, highly specialized laboratory. Although usually accurate, leukemia classification remains imperfect and errors do occur.

Distinguishing ALL from AML is critical for successful treatment; chemotherapy regimens for ALL generally contain corticosteroids, vincristine, methotrexate, and L-asparaginase, whereas most AML regimens rely on a backbone of daunorubicin and cytarabine (Pui, C.H., and Evans, W.E., *N. Engl.*

J. Med. 339:605 (1998); Bishop, J.F., *Med. J. Aust. 170*:39 (1999); Stone, R.M. and Mayer, R.J., *Hematol. Oncol. Clin. N. Am. 7*:47 (1993)). Although remission can be achieved using ALL therapy for AML (and vice versa), cure rates are markedly diminished, and unwarranted toxicities are encountered.

Recently, Golub, T.R., *et al.*, *Science 286*: 531-537 (October1999), have reported on a cancer classification scheme for AML and ALL based on the gene expression monitoring of 50 human genes. Although the 50-gene predictor approach for diagnosing AML versus ALL fared well in validation studies, the Golub *et al.* report noted that the average prediction strength was lower for samples from a different laboratory, thus emphasizing the importance of standardizing sample preparation. Further, the application of 50 genes for AML-ALL class distinction may not be desirable for a clinical setting. A method/tool employing fewer indicator genes/gene products than used by Golub *et al.* would provide increased ease, increased speed, and reduced cost. Potential for human error (misidentification) could be reduced. Reliance on expert, trained interpretation of data could also be reduced. Rapid diagnosis based on the non-random correlations ("diagnostic signatures" or "fingerprints") according to the invention described below thus would produce enormous benefit. Clearly, there is a continued need for simpler and less costly objective cancer classification approaches, especially for the classification of acute leukemias.

Summary of the Invention

The inventors have discovered that measuring the levels of small combinations (two, three or more) of particular human genes (in terms of nucleic acid or protein levels) can be used to distinguish AML from ALL. Accordingly, the present invention overcomes the disadvantages of the prior art by providing a method for diagnosing leukemia by measuring the levels of a lesser number of genes than provided in the art.

The invention also provides a preferred embodiment of the foregoing method wherein the human genes used to diagnose are LYN V-yes-1 Yamaguchi

sarcoma viral related oncogene homolog, PPGB Protective protein for beta-galactosidase, and Zyxin.

In the most preferred embodiment of the foregoing method, the genes used to diagnose are: leukotriene C4 synthase (LTC4S) gene and Zyxin.

The invention also provides a very particularly preferred embodiment of the foregoing methods, wherein the level of gene expression is measured using a DNA microchip.

The present invention also provides an embodiment, whereby the measurement of at least two human genes is used as a prognostic indicator of AML.

The present invention also provides a kit for diagnosis or prognosis of leukemia.

The invention also relates to therapies targeted at the indicator genes described herein, as well as the screening of drugs for cancer that target these indicator genes or their protein products.

It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory and are intended to provide further explanation of the invention as claimed.

Detailed Description of the Preferred Embodiments

The inventors have discovered that measurement of the levels of only a few human genes (nucleic acid levels or protein levels) can be used to distinguish AML from ALL. By "nucleic acid" is intended RNA or DNA, preferably mRNA or cDNA derived therefrom. Accordingly, the present invention overcomes the disadvantages of the prior art such as Golub *et al.* (1999), *supra*, by providing a method for diagnosing and classifying acute leukemia by measuring the expression levels of a lesser number of genes or gene products.

The names of the genes useful in diagnosis and/or prognosis described herein are as designated by Affymetrix and Golub *et al.*, and, according to them, correspond, as indicated in Appendix B, to particular GenBank entries.

The invention also provides a preferred embodiment of the foregoing method wherein the human genes used to diagnose are: LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog, PPGB Protective protein for beta-galactosidase, and Zyxin. These gene names are as assigned by Affymetrix and Golub *et al.*, and according to them, correspond to GenBank Accession Nos. M16038_at, M22960_at, and X95735_at, respectively.

In the most preferred embodiment of the foregoing method, the genes used to diagnose are: leukotriene C4 synthase (LTC4S) gene and Zyxin. These gene names are as assigned by Affymetrix and Golub *et al.*, according to them, correspond to GenBank Accession Nos. U50136_mal_at, and X95735_at, respectively.

Other embodiments employ other csets which are identified in Appendix A.

It is expected that, for certain csets, an inverse pattern of gene expression of ALL markers, as disclosed herein, would correlate with AML diagnosis. Likewise, an inverse pattern of gene expression of AML markers, as disclosed herein, would correlate with ALL diagnosis.

The invention also provides a very particularly preferred embodiment of the foregoing methods, wherein the level of gene expression is measured using a DNA microchip.

The present invention also provides an embodiment, whereby the measurement of small combinations (two, three or more) of particular human genes is used as a prognostic indicator of AML.

The present invention also provides a kit for diagnosis or prognosis of leukemia.

Gene expression data from the database http://waldo.wi.mit.edu/MPR/data_set_ALL_AML.html (which was made publicly available on October 15, 1999) was analyzed as described below. Per Golub *et al.*, *Science 286*: 531-537 (Oct 15 1999), incorporated herein by reference, the database contains the levels of expression of each of 7129 genes for each of 72 leukemia samples, which levels were determined using Affymetrix

genechip technology. The samples were classified by Golub *et al.* as either acute myeloid leukemia (AML) or acute lymphoblastic leukemia (ALL) and this information is also included in the database. The database further includes clinical data on 15 individual acute myeloid leukemia (AML) samples, with respect to treatment success or failure.

The present inventors set out to detect signal(s) from the noise in the huge data set, *i.e.*, to identify previously unrecognized correlated gene expression levels of groups of genes. To this end, the raw gene expression data was used in that form or processed using a standard data normalization technique (linear transformation followed by logarithm). Next, the expression levels for each gene were subjected to one of two standard data clustering techniques ("K means" as practiced by those skilled in the art or "Mutual nearest neighbors" as described in Jarvis, R.A. and Patrick, E.A., *IEE Trans. Computers C-22*:1025-1034 (1973)). Such pre-processing made the subsequent identification of correlations more convenient. "Clustering", as it is commonly held in the art, refers to methods for grouping "objects" of a system based on some similarity measure. The set of values in the system being analyzed is replaced by another, smaller set of values in a way that reflects the original distribution according to a chosen distance metric. In effect, clustering forces objects into likely groups. Here, the objects were the various experimentally determined levels of expression of a particular gene. The clustering algorithm provided grouping of the expression level for each gene into classes, as set forth in Appendix A. For example, referring to line 3 of Appendix A (cset 2), experimentally determined expression levels of gene 1745 may be grouped into low (A, mean = 429.4) and high (B, mean = 2211.2). In contrast, the grouping of expression levels for gene 3320, line 1 (cset1) was into three classes, low (A, mean = 923.6), medium (B, mean = 2405.8), and high (C, mean = 3496.8). (See Appendix B for the Affymetrix and Golub *et al.* assigned name corresponding to the gene numbers employed herein. For example, gene 1745 corresponds to Affymetrix and Golub *et al.* name LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog).

Next, the pre-processed data was subjected to a variant of the "coincidence detection" method described in International Patent Publication No. WO 98/43182, published October 1, 1998 (incorporated herein by reference). This method provides the identification of features which are sets of attributes (values) that co-occur more often than by random assortment and, accordingly, the identification of inherent, often unexpected features of a system. Unlike other approaches to such identification, the number of members of the identified set is not chosen prior to application of the method. That is, some approaches seek correlations between pairs of attributes (binary or 2-ary correlations). Instead, the coincidence detection method does not impose that k (as in k-ary correlations) be any specific number. Rather, the patterns inherent in the system are uncovered. As employed herein, "objects" were samples and "attributes" were gene expression values for particular genes, the ALL versus AML diagnosis , and treatment outcome for some AML samples. The high-order correlations ("coincidence sets" or "csets") discovered by the coincidence detection method were further filtered and sorted by application of another correlation test. Matthews correlation (also known as "Four-point Correlation") is a standard, known, though less commonly-used variant of the standard Pearson correlation measure, especially suited for discrete (as opposed to continuous) data. In this case, a Matthews correlation was calculated between (1) particular correlated gene expression values, considered together for the k genes in the particular cset and (2) the attribute corresponding to AML or ALL diagnosis, and the csets were sorted from highest to lowest Matthews correlation. These Matthews-tagged csets may be interpreted as "rules" relating particular genes and their expression-value ranges to diagnosis or prognosis. A plausible English interpretation of such a discovered rule (see second cset in appendix A) might be, for example,

" Gene 1745 has expression level A (LOW relative to a control, that is, value closest to the calculated cluster mean of 429 for this gene in one analysis performed and described herein) AND Gene 1829 has value B (LOW relative to a control) AND Gene 4847 has value A (LOW relative to a control) IF AND

ONLY IF the patient has leukemia type ALL (with probability based on Matthews correlation of 0.9077)."

Appendix A shows csets obtained from clustered raw data and from clustered log normalized data. Where the same cset appears more than once in Appendix A, this derives from results of multiple experimental runs (different clustering techniques).

Thus, using these techniques, the present inventors discovered small combinations of genes that provide a diagnostic indication of acute leukemia subtype. In addition, they also discovered small combinations of genes that provide a prognostic indication for AML.

As these results indicate dependence of leukemia subtype on clustered gene expression levels, they are also indicative of dependence of the subtype on unclustered (or raw) gene expression levels. This latter relationship was confirmed by the present inventors using supervised learning techniques (artificial neural networks, decision trees, etc.) as known by those skilled in the art and as described in Mitchell, T.B., *in:* Machine Learning, chapters 3 and 4, McGraw-Hill (1997). The expression levels, for the genes discovered by the coincidence detection method, were given (in raw form, that is, unnormalized and unclustered) to the supervised learning agent and the subtype of leukemia (AML versus ALL) was predicted. The training of a neural network, and the use of a trained neural network for prediction or classification, is well known to those skilled in the art.

Genes correlated with specific disease subtypes are likely to have a specific role in the disease condition, and hence are valuable targets for new therapeutics.

Genes correlated with disease prognosis are likely to have a specific role in the disease condition, and hence are valuable targets for new therapeutics. Accordingly, the invention provides methods of screening for drugs that modulate (enhance or inhibit) expression of genes in the csets, or modulate (enhance or inhibit) the activity of products of such genes.

For example, screening methods for identifying compounds capable of treating acute leukemia include contacting cells with the candidate compound, measuring gene expression, and comparing the gene expression of a particular cset to a standard expression of a particular cset, the standard being assayed when contact is made in absence of the candidate compound; whereby, a difference in gene expression indicated that the compound may be useful for treating particular subtypes of acute leukemia.

High-order correlated genes are likely to play a synergistic or antagonistic role in the disease condition, and are likely to reveal important pathways involved in the disease process.

Certain tissues in mammals with leukemia express enhanced and/or diminished levels of certain proteins and mRNA when compared to a corresponding "standard" mammal, *i.e.*, a mammal of the same species not having the leukemia. Further, it is believed that enhanced levels of certain proteins and mRNA can be detected in certain body fluids (*e.g.*, sera, plasma, urine, and spinal fluid) from mammals with leukemia when compared to body fluids from mammals of the same species not having the leukemia. Thus, the invention provides a diagnostic method useful during leukemia diagnosis, which involves assaying the expression level of a gene or set of genes in mammalian cells or body fluid and comparing the gene expression level with a standard gene expression level, whereby a difference in the gene expression level over the standard is indicative of a specific type of leukemia. In the working examples disclosed herein, comparison was made between ALL and AML samples.

Where a leukemia diagnosis has already been made according to conventional methods, the present invention is useful for confirmation thereof and as a prognostic indicator, where patients exhibiting differing gene expression will experience a better or worse clinical outcome relative to other patients.

By "assaying the level of the gene expression" is intended qualitatively or quantitatively measuring or estimating the level of the protein or the level of the mRNA encoding the protein in a first biological sample either directly (*e.g.*, by determining or estimating absolute protein level or mRNA level) or relatively

(*e.g.*, by comparing to the protein level or mRNA level in a second biological sample).

Preferably, the protein level or mRNA level in the first biological sample is measured or estimated and compared to a standard protein level or mRNA level (*e.g.*, ALL sample v. AML sample), the standard being taken from a second biological sample obtained from an individual not having that leukemia. As will be appreciated in the art, once a standard protein level or mRNA level is known, it can be used repeatedly as a standard for comparison.

By "biological sample" is intended any biological sample obtained from an individual, cell line, tissue culture, or other source which contains protein or mRNA. Biological samples include mammalian body fluids (such as sera, plasma, urine, synovial fluid and spinal fluid) which contain secreted mature protein, and ovarian, prostate, heart, placenta, pancreas liver, spleen, lung, breast and umbilical tissue.

The present invention is useful for detecting acute leukemia in mammals. Preferred mammals include monkeys, apes, cats, dogs, cows, pigs, horses, rabbits and humans. Particularly preferred are humans.

In order to detect gene expression, total cellular RNA can be isolated from a biological sample using the single-step guanidinium-thiocyanate-phenol-chloroform method described in Chomczynski and Sacchi, *Anal. Biochem. 162:*156-159 (1987). Levels of mRNA encoding the protein (or cDNA prepared from such mRNA) are then assayed using any appropriate method. These include Northern blot analysis (Harada *et al., Cell 63:*303-312 (1990)), S1 nuclease mapping (Fujita *et al., Cell 49:*357- 367 (1987)), the polymerase chain reaction (PCR), reverse transcription in combination with the polymerase chain reaction (RT-PCR) (Makino *et al., Technique 2:*295-301 (1990)), and reverse transcription in combination with the ligase chain reaction (RT-LCR).

Protein levels may be determined by assaying enzymatic activity of the protein. This is especially useful when screening potentially useful therapeutic drugs that affect protein activity.

Assaying protein levels in a biological sample can also be performed using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods (Jalkanen, M., *et al., J. Cell. Biol. 101*:976-985 (1985); Jalkanen, M., *et al., J. Cell. Biol. 105*:3087-3096 (1987)). This is useful when screening drugs as potential therapeutics that affect gene expression.

Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA).

Suitable labels are known in the art and include enzyme labels, such as, glucose oxidase, horseradish peroxidase and alkaline phosphatase; radioisotopes, such as iodine (^{125}I, ^{121}I), carbon (^{14}C), sulfur (^{35}S), tritium (^{3}H), indium (^{112}In), and technetium (^{99m}Tc); fluorescent labels, such as fluorescein and rhodamine; and biotin.

In a preferred embodiment, gene expression is measured using a DNA microchip, as described below in Example 3. DNA microchips are described in U.S. Patent Nos. 5,744,305; 5,424,186; 5,412,087; 5,489,678; 5,889,165; 5,753,788; and 5,744,101; and WO 98/12559; and Harris, *Exp. Opin. Ther. Patents 5*:469-476 (1995). DNA microchips contain oligonucleotide probes affixed to a solid substrate, and are useful for screening a large number of samples for gene expression.

The present invention also further includes kits for diagnosing subtypes of acute leukemia, comprising a means for measuring gene expression of each gene of a cset which is herein disclosed as being correlated with a subtype of leukemia, wherein said means are within a container. In one embodiment, a kit is provided which comprises a means for measuring gene expression of LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog, a means for measuring gene expression of PPGB Protective protein for beta-galactosidase, and a means for measuring gene expression of Zyxin. In one embodiment, the means for measuring gene expression is a DNA microchip which contains probes specific for the target gene(s). In another embodiment, the means for measuring

gene expression is an antibody specific for the protein of interest. Other means for measuring gene expression are well known in the art.

The invention also relates to therapies targeted at these indicator genes, as well as the screening of drugs for cancer that target these indicator genes or their protein products.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

Examples

Example 1

Those skilled in the art can, by the exercise of ordinary skill, measure the mRNA or protein level for each of the two, three or more (preferably two to six) genes in a correlated set discovered to be diagnostic for leukemia subtype and, in reference to a standard, classify new cases of leukemia with respect to subtype. Such an analysis would be highly amenable to modern diagnostic "chip" technology and suitable for incorporation into a bedside diagnostic device.

For example, in reference to Appendix A, page a, cset 2, the expression level of Affymetrix designated genes LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog (GenBank Accession #M16038), PPGB Protective protein for beta-galactosidase (galactosialidosis) (GenBank Accession #M22960), and Zyxin (GenBank Accession #X95735) is diagnostic of ALL. In this case, diagnosis of ALL can be made if the relative expression level of each of these genes is low. Similarly, other csets in Appendix A provide diagnostic gene "signatures" or "fingerprints" of similar value.

Example 2

Those skilled in the art can measure the mRNA or protein level for each of the genes in a correlated set discovered to be a prognostic indicator for AML,

and in reference to a standard, predict patient response to treatment. Such an analysis could be extremely valuable in designating patients as unlikely to respond to conventional therapy, and hence targeting them for more intensive or more experimental procedures.

For example, in reference to Appendix C, cset 2, the expression level of genes 1436 and 3847 (Affymetrix designated genes POU3F1 POU domain, class 3, transcription factor 1, GenBank Accession No. L26494_at; and GB DEF = homeodomain protein HoxA9 mRNA, GenBank Accession No. U82759_at, respectively) is a prognostic indicator for AML. In this case, AML prognosis is good if the relative expression level of these genes is medium-high and high, respectively.

Example 3

Total RNA is extracted from tissue samples of a patient with leukemia, and cDNA is prepared using methods well known in the art. Double-stranded DNA is made from the cDNA. The double-stranded cDNA is transcribed using the Ambion T7 MegaScript Kit. The cRNA made from the in vitro-translation of the double-stranded cDNA is fragmented by adding 15 µg cRNA to 0.2 vol of 5X fragmentation buffer and storing at 95°C for 35 minutes. The fragmented cRNA is then added to 3 uL 5 nM Control Oligonucleotide B2 (Final concentration: 50 pM)(Affymetrix); 3 uL 10 mg/ml Herring Sperm DNA (Final concentration: 0.1 mg/ml)(Promega/Fisher Scientific); 3 uL 50 mg/ml Acetylated BSA (Final concentration: 0.5 mg/ml)(Gibco BRL Life Technologies); 150 ul 2X MES Hybridization Buffer (Final concentration: 1X). The volume is adjusted with DEPC H_20 to 300 uL total volume.

A 12X MES Stock buffer is prepared: 70.4 g MES free acid monohydrate (Final concentration: 1.22 M MES)(Sigma Chemicals); 193.3 g MES sodium salt (Final concentration: 0.89M [Na+])(Sigma Chemicals); 800 ml DEPC H_2O; the volume is brought up with water to 1000 ml. pH should be between 6.5 and 6.7.

A DNA microchip, containing probes for LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog, PPGB Protective protein for beta-galactosidase, and Zyxin, is prepared using, for example, the methods described in U.S. Patent No. 5,744,305, which is herein incorporated by reference. The microchip is equilibrated to room temperature just before use. The chips are pre-wet with 200 uL of 1X MES Hybridization buffer at 45°C for 10-20 minutes, 60 RPM. The fragmented cRNA is heated at 99°C for 5 minutes and cooled at 45°C for 5 minutes, then spun at maximum speed for 5 minutes. The 1X MES hybridization buffer is removed from chips, and 200 µl fragmented cRNA is added to each chip. The chips are incubated at 45°C, 60 RPM for 16 hours. After 16 hour hybridization, the cRNA is removed from the chip and stored at -80°C.

For each chip: 1200 uL SAPE (Streptavidin Phycoerythrin) Solution is prepared, using 600 uL 2X Stain buffer; 120 uL 20 mg/mL Acetylated BSA (Final concentration: 2 mg/mL); 12 uL 1 mg/mL SAPE (Final Concentration: 10 ug/mL)(Molecular Probes); 468 uL DEPC H_2O. 600 uL Antibody Solution is prepared, using: 300 uL 2X Stain Buffer; 60 uL 20 mg/mL Acetylated BSA (Final concentration: 2mg/mL); 30 uL goat serum (Final concentration: 5%)(Sigma Chemical); 3.6 uL 0.5 mg/mL biotinylated anti-streptavidin antibody (Final concentration: 3 ug/mL)(Vector Laboratories); and 206.4 uL DEPC H_2O.

2X Stain buffer is prepared using 41.7 ml 12X MES Stock Buffer (Final concentration: 100 mM MES); 92.5 ml 5 M NaCl (Final concentration: 1 M [Na+]); 2.5 ml 10% Tween 20 (Final concentration: 0.05% Tween); 112.8 ml DEPC H_2O; filtering through a 0.2 um filter; after filtering, add 0.5 ml of 5% Antifoam.

Hybridization is performed using the Affymetrix GeneChip© Fluidics Station 400 at 10 cycles of 2 mixes per cycle with Non-Stringent Wash Buffer at 25°C; 4 cycles of 15 mixes per cycle with Stringent Wash Buffer at 50°C; probe is stained with the first aliquot of the SAPE solution for 10 minutes at 25°C; 10 cycles of 4 mixes per cycle at 2°C; probe is stained in antibody solution for 10 minutes at 25°C; probe is stained with the second aliquot of SAPE for 10 minutes

at 25°C; final wash is 15 cycles of 4 mixes per cycles at 30°C; holds at 25°C. The plates are scanned using the Hewlett-Packard GeneArray© Scanner (Affymetrix).

Example 4

Those skilled in the art can, by the exercise of ordinary skill, measure the mRNA or protein level for each of the two, three or more (preferably two to six) in a correlated set discovered to be diagnostic for leukemia subtype and, in reference to a standard, classify new cases of leukemia with respect to subtype. Such an analysis would be highly amenable to modern diagnostic "chip" technology and suitable for incorporation into a bedside diagnostic device.

For example, in reference to Appendix A, page i, cset 1 for AML, the expression level of Affymetrix designated genes Zyxin (GenBank Accession #X95735_at) and ELA2 Elastase 2, neutrophil (GenBank Accession #M27783_at) is diagnostic of AML. In this case, diagnosis of AML can be made if the relative expression level of each of these genes is high. Similarly, other csets in Appendix A provide diagnostic gene "signatures" or "fingerprints" of similar value.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described.

The entire disclosure of all publications (including patents, patent applications, journal articles, databases, GenBank entries, web sites, laboratory manuals, books, or other documents) cited herein are hereby incorporated by reference.

Appendix A

ALL Predictors
Clustered Raw Data

Matthews Relation	Observed Association	
0.9094	45ALL	Value:C Gene:3320 where A=2405.82 B=3498.8 C=923.571
0.9077	46ALL	Value:A Gene:4847 where A=318.787 B=3397.48
		Value:A Gene:1745 where A=429.413793 B=2211.214286
		Value:B Gene:1829 where A=2450.666667 B=522.245614
0.8813	44ALL	Value:A Gene:4847 where A=434.117647 B=3703.809524
		Value:A Gene:2288 where A=28.181818 B=7065.235294
		Value:A Gene:3252 where A=101.470588 B=1662.000000
		Value:B Gene:3320 where A=2693.235294 B=906.963636
0.8774	45ALL	Value:A Gene:4847 where A=434.117647 B=3703.809524
		Value:C Gene:760 where A=8172.4 B=3964 C=376.25
0.8774	45ALL	Value:A Gene:4847 where A=318.787 B=3397.48
0.8768	46ALL	Value:A Gene:4847 where A=318.787 B=3397.48
		Value:A Gene:4847 where A=434.117647 B=3703.809524
0.8768	48ALL	Value:A Gene:6919 where A=280.034483 B=1432.428571
		Value:A Gene:1779 where A=884.650000 B=15238.916667
0.8768	46ALL	Value:A Gene:4847 where A=434.117647 B=3703.809524
		Value:A Gene:2288 where A=28.181818 B=7065.235294
0.8629	42ALL	Value:A Gene:4847 where A=434.117647 B=3703.809524
		Value:A Gene:2121 where A=1739.65 B=6935.94
0.8629	42ALL	Value:A Gene:3252 where A=52.0476 B=1536.46 C=169.333
0.8629	42ALL	Value:A Gene:3252 where A=52.0476 B=1536.46 C=169.333
		Value:A Gene:3252 where A=52.0476 B=1536.46 C=169.333
0.8486	44ALL	Value:C Gene:3320 where A=2405.82 B=3498.8 C=923.571
		Value:A Gene:1779 where A=884.650000 B=15238.916667
		Value:A Gene:3252 where A=101.470588 B=1662.000000
0.8486	44ALL	Value:A Gene:4847 where A=434.117647 B=3703.809524
		Value:A Gene:1882 where A=770.250000 B=15876.000000
		Value:A Gene:2288 where A=28.181818 B=7065.235294
		Value:A Gene:3252 where A=101.470588 B=1662.000000
		Value:A Gene:4847 where A=434.117647 B=3703.809524
0.8486	44ALL	Value:A Gene:6376 where A=186.475410 B=2425.818182
		Value:A Gene:2288 where A=28.181818 B=7065.235294
		Value:A Gene:3252 where A=101.470588 B=1662.000000
0.8486	44ALL	Value:A Gene:4847 where A=434.117647 B=3703.809524
		Value:A Gene:3252 where A=101.470588 B=1662.000000
		Value:A Gene:4847 where A=434.117647 B=3703.809524
0.8462	48ALL	Value:A Gene:2121 where A=1739.65 B=6935.94
0.8458	45ALL	Value:C Gene:3320 where A=2405.82 B=3498.8 C=923.571
		Value:B Gene:1829 where A=2450.666667 B=522.245614
		Value:A Gene:3252 where A=101.470588 B=1662.000000
0.8458	45ALL	Value:B Gene:3320 where A=2693.235294 B=906.963636
		Value:A Gene:2288 where A=28.181818 B=7065.235294
		Value:A Gene:3252 where A=101.470588 B=1662.000000

Matthews Relation	Observed Ass ciation	
		Value:A Gene:6803 where A=2025.766885 B=10902.181818
		Value:A Gene:6806 where A=1858.393443 B=10826.818182
0.8387	41ALL	Value:A Gene:804 where A=3301.48 B=10857 C=692.615
		Value:A Gene:3252 where A=52.0476 B=1536.46 C=169.333
0.8210	43ALL	Value:A Gene:2242 where A=44.150000 B=538.750000
		Value:B Gene:3847 where A=887.588235 B=182.090909
		Value:A Gene:4847 where A=434.117647 B=3703.809524
0.8210	43ALL	Value:B Gene:1829 where A=2450.666667 B=522.245614
		Value:A Gene:1834 where A=234.559322 B=1245.538462
		Value:A Gene:3252 where A=101.470588 B=1662.000000
		Value:B Gene:3320 where A=2693.235294 B=906.963636
		Value:B Gene:4499 where A=972.454545 B=209.032787
		Value:A Gene:5663 where A=778.763636 B=2486.647059
0.8157	46ALL	Value:A Gene:4847 where A=434.117647 B=3703.809524
0.8154	40ALL	Value:A Gene:2121 where A=1739.65 B=6935.94
		Value:A Gene:3252 where A=52.0476 B=1536.46 C=169.333
		Value:A Gene:4847 where A=318.787 B=3397.48
		Value:B Gene:2128 where A=576.2 B=292.891 C=1277.12
0.8154	40ALL	D=7459
		Value:A Gene:4847 where A=318.787 B=3397.48
0.8154	40ALL	Value:A Gene:2383 where A=522.293 B=2712
		Value:A Gene:3252 where A=52.0476 B=1536.46 C=169.333
		Value:A Gene:4847 where A=318.787 B=3397.48
0.8154	40ALL	Value:A Gene:3252 where A=52.0476 B=1536.46 C=169.333
		Value:A Gene:4847 where A=318.787 B=3397.48
0.8143	45ALL	Value:A Gene:804 where A=3301.48 B=10857 C=692.615
		Value:A Gene:2121 where A=1739.65 B=6935.94
0.8143	45ALL	Value:A Gene:4847 where A=434.117647 B=3703.809524
		Value:A Gene:6201 where A=890.474576 B=13711.461538
0.8143	45ALL	Value:A Gene:4847 where A=434.117647 B=3703.809524
		Value:A Gene:6041 where A=651.929825 B=3705.800000
0.8143	45ALL	Value:B Gene:1829 where A=2450.666667 B=522.245614
		Value:A Gene:1834 where A=234.559322 B=1245.538462
		Value:A Gene:3252 where A=101.470588 B=1662.000000
0.8143	45ALL	Value:A Gene:4366 where A=343.290909 B=2419.882353
		Value:A Gene:4847 where A=434.117647 B=3703.809524
0.8038	41ALL	Value:A Gene:1834 where A=234.559322 B=1245.538462
		Value:A Gene:2121 where A=1946.135593 B=7997.384615
		Value:A Gene:2288 where A=28.181818 B=7065.235294
		Value:A Gene:3482 where A=-37.711864 B=67.384615
		Value:A Gene:4196 where A=1409.291667 B=7309.875000
		Value:A Gene:4847 where A=434.117647 B=3703.809524

ALL Predictors
Clustered Log Normalized Data

Matthews Relation	Observed	Ass ciation	
0.9095	47	ALL	Value:A Gene:1779 where A=4.466110 B=4.634806
			Value:A Gene:1882 where A=4.462936 B=4.637279
			Value:A Gene:2121 where A=4.481919 B=4.560041
			Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:2402 where A=4.467723 B=4.633272
			Value:A Gene:6376 where A=4.455837 B=4.488488
0.8813	44	ALL	Value:A Gene:1615 where A=4.462970 B=4.488003
			Value:A Gene:3482 where A=4.452756 B=4.454362
			Value:A Gene:4847 where A=4.458925 B=4.504069
0.8813	44	ALL	Value:B Gene:3320 where A=4.492605 B=4.466959
			Value:A Gene:4847 where A=4.458925 B=4.504069
0.8813	44	ALL	Value:A Gene:1745 where A=4.459603 B=4.484955
			Value:B Gene:3320 where A=4.492605 B=4.466959
			Value:A Gene:4847 where A=4.458925 B=4.504069
0.8774	45	ALL	Value:A Gene:1745 where A=4.459603 B=4.484955
			Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:3252 where A=4.478901 B=4.548814
			Value:A Gene:4847 where A=4.458925 B=4.504069
0.8774	45	ALL	Value:A Gene:1745 where A=4.459603 B=4.484955
			Value:B Gene:4499 where A=4.467696 B=4.456513
			Value:A Gene:4847 where A=4.458925 B=4.504069
0.8774	45	ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:3252 where A=4.454877 B=4.477933
			Value:B Gene:3320 where A=4.492605 B=4.466959
0.8544	43	ALL	Value:A Gene:1779 where A=4.466110 B=4.634806
			Value:A Gene:3252 where A=4.454877 B=4.477933
			Value:A Gene:4190 where A=4.453190 B=4.476172
			Value:A Gene:5432 where A=4.453427 B=4.455339
			Value:A Gene:6201 where A=4.463253 B=4.612053
0.8544	43	ALL	Value:A Gene:3252 where A=4.454877 B=4.477933
			Value:B Gene:3320 where A=4.492605 B=4.466959
			Value:A Gene:4229 where A=4.453830 B=4.469877
			Value:A Gene:4847 where A=4.458925 B=4.504069
			Value:A Gene:6563 where A=4.462051 B=4.490602
0.8503	47	ALL	Value:A Gene:1834 where A=4.456900 B=4.471925
			Value:A Gene:2121 where A=4.481919 B=4.560041
			Value:A Gene:2288 where A=4.453667 B=4.547863
0.8503	47	ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:B Gene:4499 where A=4.467696 B=4.456513
0.8486	44	ALL	Value:B Gene:1829 where A=4.488279 B=4.460958
			Value:A Gene:4847 where A=4.458925 B=4.504069
0.8486	44	ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:3252 where A=4.454877 B=4.477933
			Value:B Gene:3320 where A=4.492605 B=4.466959
			Value:A Gene:5833 where A=4.450558 B=4.463545
0.8486	44	ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:4847 where A=4.458925 B=4.504069

Matthews Relation	Observed Association	
		Value:A Gene:6201 where A=4.483253 B=4.612053
0.8462	46ALL	Value:A Gene:1834 where A=4.456900 B=4.471925
		Value:A Gene:1882 where A=4.462936 B=4.637279
		Value:B Gene:3320 where A=4.492605 B=4.466959
		Value:A Gene:6803 where A=4.481823 B=4.588492
		Value:A Gene:6806 where A=4.479556 B=4.587092
0.8462	46ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:B Gene:3320 where A=4.492605 B=4.466959
0.8462	46ALL	Value:B Gene:1829 where A=4.488279 B=4.460958
		Value:A Gene:1834 where A=4.456900 B=4.471925
		Value:A Gene:2288 where A=4.453667 B=4.547863
0.8458	45ALL	Value:B Gene:1829 where A=4.488279 B=4.460958
		Value:A Gene:1834 where A=4.456900 B=4.471925
		Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:A Gene:5833 where A=4.450558 B=4.463545
		Value:A Gene:6919 where A=4.457584 B=4.474643
0.8458	45ALL	Value:B Gene:1829 where A=4.488279 B=4.460958
		Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:A Gene:6185 where A=4.465723 B=4.524227
0.8458	45ALL	Value:A Gene:1882 where A=4.462936 B=4.637279
		Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:A Gene:2585 where A=4.455314 B=4.463555
		Value:A Gene:3252 where A=4.454877 B=4.477933
		Value:A Gene:4229 where A=4.453830 B=4.489877
		Value:A Gene:6797 where A=4.482005 B=4.586722
		Value:A Gene:6803 where A=4.481823 B=4.588492
		Value:A Gene:6806 where A=4.479556 B=4.587092
		Value:A Gene:6919 where A=4.457584 B=4.474643
0.8458	45ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:A Gene:2402 where A=4.467723 B=4.633272
		Value:A Gene:4847 where A=4.458925 B=4.504069
0.8458	45ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:A Gene:4847 where A=4.458925 B=4.504069
0.8458	45ALL	Value:A Gene:4847 where A=4.458925 B=4.504069
		Value:A Gene:6919 where A=4.457584 B=4.474643
0.8458	45ALL	Value:A Gene:1882 where A=4.462936 B=4.637279
		Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:A Gene:3252 where A=4.454877 B=4.477933
0.8458	45ALL	Value:A Gene:1882 where A=4.462936 B=4.637279
		Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:A Gene:3252 where A=4.454877 B=4.477933
		Value:A Gene:6803 where A=4.481823 B=4.588492
		Value:A Gene:6806 where A=4.479556 B=4.587092
0.8458	45ALL	Value:A Gene:1882 where A=4.462936 B=4.637279
		Value:A Gene:4847 where A=4.458925 B=4.504069
0.8458	45ALL	Value:A Gene:1882 where A=4.462936 B=4.637279
		Value:A Gene:4847 where A=4.458925 B=4.504069
		Value:A Gene:6797 where A=4.482005 B=4.586722
		Value:A Gene:6803 where A=4.481823 B=4.588492
		Value:A Gene:6806 where A=4.479556 B=4.587092

Matthews Relation	Observed Association	
0.8458	45ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:A Gene:4847 where A=4.458925 B=4.504069
0.8458	45ALL	Value:A Gene:6919 where A=4.457584 B=4.474643
		Value:A Gene:3252 where A=4.454877 B=4.477933
0.8458	45ALL	Value:B Gene:3320 where A=4.492605 B=4.466959
		Value:A Gene:4847 where A=4.458925 B=4.504069
0.8286	42ALL	Value:A Gene:6797 where A=4.482005 B=4.586722
		Value:A Gene:6803 where A=4.481823 B=4.588492
		Value:A Gene:6806 where A=4.479556 B=4.587092
		Value:B Gene:1829 where A=4.488279 B=4.460958
		Value:A Gene:1834 where A=4.456900 B=4.471925
		Value:A Gene:3183 where A=4.480033 B=4.507884
0.8210	43ALL	Value:B Gene:3320 where A=4.492605 B=4.466959
		Value:A Gene:4377 where A=4.463571 B=4.492668
		Value:A Gene:2288 where A=4.453667 B=4.547863
0.8210	43ALL	Value:A Gene:3252 where A=4.454877 B=4.477933
		Value:A Gene:4847 where A=4.458925 B=4.504069
0.8210	43ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:A Gene:3252 where A=4.454877 B=4.477933
		Value:A Gene:4847 where A=4.458925 B=4.504069
0.8210	43ALL	Value:A Gene:6041 where A=4.463123 B=4.506267
		Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:A Gene:3252 where A=4.454877 B=4.477933
0.8210	43ALL	Value:A Gene:4847 where A=4.458925 B=4.504069
		Value:A Gene:6919 where A=4.457584 B=4.474643
		Value:A Gene:2363 where A=4.460906 B=4.491798
0.8210	43ALL	Value:A Gene:3252 where A=4.454877 B=4.477933
		Value:A Gene:4366 where A=4.458506 B=4.488684
		Value:A Gene:4847 where A=4.458925 B=4.504069
0.8210	43ALL	Value:A Gene:3252 where A=4.454877 B=4.477933
		Value:A Gene:4366 where A=4.458506 B=4.488684
0.8210	43ALL	Value:A Gene:4847 where A=4.458925 B=4.504069
		Value:A Gene:3252 where A=4.454877 B=4.477933
0.8162	44ALL	Value:A Gene:4847 where A=4.458925 B=4.504069
		Value:A Gene:1745 where A=4.459603 B=4.484955
		Value:A Gene:2288 where A=4.453667 B=4.547863
		Value:A Gene:4229 where A=4.453830 B=4.489877
0.8162	44ALL	Value:B Gene:4499 where A=4.467896 B=4.456513
		Value:A Gene:5280 where A=4.461424 B=4.490763
		Value:A Gene:1745 where A=4.459603 B=4.484955
		Value:A Gene:2288 where A=4.453667 B=4.547863
0.8162	44ALL	Value:A Gene:4847 where A=4.458925 B=4.504069
		Value:A Gene:5833 where A=4.450558 B=4.463545
		Value:A Gene:6919 where A=4.457584 B=4.474643
		Value:A Gene:3252 where A=4.454877 B=4.477933
		Value:A Gene:6005 where A=4.462943 B=4.480720
		Value:A Gene:6803 where A=4.481823 B=4.588492
		Value:A Gene:6806 where A=4.479556 B=4.587092
0.8162	44ALL	Value:A Gene:6919 where A=4.457584 B=4.474643
		Value:B Gene:1260 where A=4.457739 B=4.454284

Matthews Relati n	Observed	Associati n	
			Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:4847 where A=4.458925 B=4.504069
0.8162	44	ALL	Value:A Gene:1615 where A=4.462970 B=4.488003
			Value:A Gene:4847 where A=4.458925 B=4.504069
0.8162	44	ALL	Value:B Gene:1829 where A=4.468279 B=4.480958
			Value:A Gene:2242 where A=4.454008 B=4.461486
			Value:A Gene:6201 where A=4.463253 B=4.612053
			Value:A Gene:6584 where A=4.458951 B=4.474055
0.8162	44	ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:3252 where A=4.454877 B=4.477933
			Value:A Gene:5833 where A=4.460558 B=4.463545
			Value:A Gene:6041 where A=4.463123 B=4.506267
0.8162	44	ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:4847 where A=4.458925 B=4.504069
			Value:A Gene:5833 where A=4.460558 B=4.463545
0.8162	44	ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:4847 where A=4.458925 B=4.504069
			Value:A Gene:5833 where A=4.460558 B=4.463545
			Value:A Gene:6919 where A=4.457584 B=4.474643
0.8162	44	ALL	Value:A Gene:4847 where A=4.458925 B=4.504069
			Value:A Gene:6185 where A=4.465723 B=4.524227
			Value:A Gene:6919 where A=4.457584 B=4.474643
0.8162	44	ALL	Value:A Gene:4847 where A=4.458925 B=4.504069
			Value:A Gene:6201 where A=4.463253 B=4.612053
0.8157	46	ALL	Value:A Gene:1779 where A=4.466110 B=4.634806
			Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:5833 where A=4.460558 B=4.463545
0.8143	45	ALL	Value:A Gene:4847 where A=4.458925 B=4.504069
0.8143	45	ALL	Value:B Gene:1260 where A=4.467739 B=4.484284
			Value:A Gene:1400 where A=4.470168 B=4.545801
			Value:A Gene:2137 where A=4.451155 B=4.461070
			Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:4366 where A=4.458506 B=4.486684
			Value:A Gene:6041 where A=4.463123 B=4.506267
0.8143	45	ALL	Value:A Gene:1745 where A=4.459603 B=4.484955
			Value:A Gene:4847 where A=4.458925 B=4.504069
0.8143	45	ALL	Value:A Gene:2121 where A=4.481919 B=4.560041
			Value:A Gene:4847 where A=4.458925 B=4.504069
0.8143	45	ALL	Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:3252 where A=4.454877 B=4.477933
0.8038	41	ALL	Value:B Gene:997 where A=4.455181 B=4.451743
			Value:A Gene:3252 where A=4.454877 B=4.477933
0.8038	41	ALL	Value:A Gene:2111 where A=4.471958 B=4.500436
			Value:A Gene:3252 where A=4.454877 B=4.477933
0.8038	41	ALL	Value:A Gene:2121 where A=4.481919 B=4.560041
			Value:A Gene:2288 where A=4.453667 B=4.547863
			Value:A Gene:4847 where A=4.458925 B=4.504069
			Value:A Gene:5107 where A=4.453589 B=4.455868
0.8038	41	ALL	Value:B Gene:997 where A=4.455181 B=4.451743
			Value:A Gene:4847 where A=4.458925 B=4.504069

Matthews R lati n	Observed Associati n	
0.8038	41ALL	Value:B Gene:1539 where A=4.458916 B=4.464273
		Value:A Gene:3258 where A=4.478301 B=4.548614
		Value:A Gene:4847 where A=4.458925 B=4.504089
0.8038	41ALL	Value:A Gene:1745 where A=4.459603 B=4.484955
		Value:A Gene:2546 where A=4.469232 B=4.499254
		Value:A Gene:3252 where A=4.454677 B=4.477933
		Value:B Gene:4499 where A=4.467898 B=4.456513
		Value:B Gene:6141 where A=4.473292 B=4.460415
		Value:B Gene:6373 where A=4.481622 B=4.461275

AML Predi tors
Clustered Raw Data

Matthews Relation	Observed	Association
0.9095	22AML	Value:B Gene:4847 where A=318.787 B=3397.48 Value:D Gene:6218 where A=7.38462 B=-157.5 C=136.158 D=4362.71 E=43
0.8798	21AML	Value:B Gene:3252 where A=52.0476 B=1536.46 C=169.333 Value:B Gene:4847 where A=318.787 B=3397.48
0.8774	23AML	Value:B Gene:4847 where A=318.787 B=3397.48 Value:B Gene:4328 where A=4603.05 B=1128.16 C=99
0.8768	22AML	D=10585 Value:B Gene:4847 where A=318.787 B=3397.48
0.8503	20AML	Value:C Gene:1144 where A=883.6 B=2760 C=238.463 Value:E Gene:2288 where A=119.125 B=-590 C=-161.634 D=19568 E=5447.25
0.8503	20AML	Value:B Gene:3252 where A=52.0476 B=1536.46 C=169.333 Value:B Gene:4847 where A=318.787 B=3397.48 Value:D Gene:1725 where A=-116 B=16.439 C=1214 D=250.207
0.8503	20AML	Value:B Gene:4328 where A=4603.05 B=1128.16 C=99 D=10585 Value:B Gene:4847 where A=318.787 B=3397.48
0.8503	20AML	Value:E Gene:2288 where A=119.125 B=-590 C=-161.634 D=19568 E=5447.25
0.8503	20AML	Value:E Gene:2288 where A=119.125 B=-590 C=-161.634 D=19568 E=5447.25 Value:B Gene:3252 where A=52.0476 B=1536.46 C=169.333
0.8503	20AML	Value:E Gene:2288 where A=119.125 B=-590 C=-161.634 D=19568 E=5447.25 Value:B Gene:3252 where A=52.0476 B=1536.46 C=169.333
0.8503	20AML	Value:B Gene:4847 where A=318.787 B=3397.48 Value:E Gene:2288 where A=119.125 B=-590 C=-161.634 D=19568 E=5447.25
0.8503	20AML	Value:B Gene:4847 where A=318.787 B=3397.48 Value:B Gene:3252 where A=52.0476 B=1536.46 C=169.333 Value:B Gene:4328 where A=4603.05 B=1128.16 C=99 D=10585
0.8462	21AML	Value:B Gene:4847 where A=318.787 B=3397.48 Value:A Gene:758 where A=84 B=1467.36 C=4015.33 D=337.5 E=7997.44 F=-65.6667
0.8462	21AML	Value:B Gene:4847 where A=318.787 B=3397.48 Value:B Gene:3252 where A=52.0476 B=1536.46 C=169.333 Value:B Gene:4328 where A=4603.05 B=1128.16 C=99 D=10585
0.8458	22AML	Value:C Gene:1902 where A=2046 B=225.6 C=-69.2821 Value:B Gene:3252 where A=52.0476 B=1536.46 C=169.333
0.8458	22AML	Value:B Gene:3252 where A=52.0476 B=1536.46 C=169.333 Value:A Gene:4196 where A=6549.6 B=1109.4
0.8458	22AML	Value:B Gene:4328 where A=4603.05 B=1128.16 C=99 D=10585
0.8210	19AML	Value:E Gene:1779 where A=-74.5 B=-257 C=1043.27 D=212.583 E=10030.4
0.8157	20AML	Value:B Gene:3252 where A=52.0476 B=1536.46 C=169.333 Value:D Gene:1725 where A=-116 B=16.439 C=1214 D=250.207

Matthews Relation	Observed Association	
0.8157	20AML	Value:B Gene:4847 where A=316.787 B=3397.48 Value:B Gene:4847 where A=434.117647 B=3703.809524

AML Predictors

Clustered Log Normalized Data

Matthews Relation	Observed Association	
0.8143	21AML	Value:B Gene:4847 where A=4.458925 B=4.504069

A281-31eppA.doc

Appendix B

Gene Index	Genbank or Affymetrix Accession Number with Submission Date	Gene Description	Reference
400	D38548 17-OCT-1994	KIAA0076 gene	Nomura,N. *et al.* 1994. Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res.* 1, 223-229 (1994)
720	D87449 27-AUG-1996	KIAA0260 gene, partial cds	Nagase,T. *et al.* 1996. Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain. *DNA Res.* 3, 321-329.
758	D88270 02-OCT-1996	DNA for immunoglobin lambda light chain	Kawasaki,K. *et al.* 1997. One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *Genome Res.* 7, 250-261 (1997)
760	D88422 15-OCT-1996	CYSTATIN A	Yamazaki,M. *et al.* 1997. Genomic structure of human cystatin A. *DNA Seq.* 8, 71-76.
804	HG1612-HT1612_at	MacMARKS	Affymetrix, Santa Clara CA
997	HG4321-HT4591_at	Ahnak-Related Sequence	Affymetrix, Santa Clara CA
1144	J05243 12-DEC-1989	SPTAN1 Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	Moon,R.T. and McMahon,A.P. 1990. Generation of diversity in nonerythroid spectrins. Multiple polypeptides are predicted by sequence analysis of cDNAs encompassing the coding region of human nonerythroid alpha-spectrin. *J. Biol. Chem.* 265, 4427-4433.
1260	L09717	LAMP2 Lysosome-associated membrane protein 2 (alternative products)	Fukuda,M. *et al.* 1988. Cloning of cDNAs encoding human lysosomal membrane glycoproteins, h-lamp-1 and h-lamp-2. Comparison of their deduced amino acid sequences. *J. Biol. Chem.* 263, 18920-18928. Sawada,R. *et al.* 1993. The genes of major lysosomal membrane glycoproteins, lamp-1 and lamp-2. 5'-flanking sequence of lamp-2 gene and comparison of exon organization in two genes. *J. Biol. Chem.* 268, 9014-9022. Erratum: *J Biol Chem* 268, 13010.
1385	L20348	Oncomodulin gene	Fohr,U.G. *et al.* 1993. Human alpha and beta parvalbumins. Structure and tissue-specific expression. *Eur. J. Biochem.* 218, 719-727.

Gene Index	Genbank or Affymetrix Accession Number with Submission Date	Gene Description	Reference
1400	L21954	PERIPHERAL-TYPE BENZODIAZEPINE RECEPTOR	Lin,D. *et al.* 1993. The human peripheral benzodiazepine receptor gene: cloning and characterization of alternative splicing in normal tissues and in a patient with congenital lipoid adrenal hyperplasia. *Genomics* 18, 643-650.
1436	L26494	POU3F1 POU domain, class 3, transcription factor 1	Faus,I., Hsu,H.J. and Fuchs,E. 1994. Oct-6: a regulator of keratinocyte gene expression in stratified squamous epithelia. *Mol. Cell. Biol.* 14, 3263-3275.
1539	L38608	ALCAM Activated leucocyte cell adhesion molecule	Bowen,M.A. *et al.*1995. Cloning, mapping, and characterization of activated leukocyte-cell adhesion molecule (ALCAM), a CD6 ligand. *J. Exp. Med.* 181, 2213-2220.
1615	L42379	Quiescin (Q6) mRNA, partial cds	Gao,C. *et al.*Molecular cloning and expression of A novel bone-derived growth factor from a human osteosarcoma cell line. Unpublished
1725	M14636	PYGL Glycogen phosphorylase L (liver form)	Newgard,C.B. *et al.* (1986) Sequence analysis of the cDNA encoding human liver glycogen phosphorylase reveals tissue-specific codon usage. *Proc. Natl. Acad. Sci. U.S.A.* 83, 8132-8136.
1745	M16038	LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog	Yamanashi,Y. *et al.* (1987) The yes-related cellular gene lyn encodes a possible tyrosine kinase similar to p56lck. *Mol. Cell. Biol.* 7, 237-243.
1779	M19507 23-NOV-1987 11-MAY-1988	MPO Myeloperoxidase	Yamada,M. *et al.* (1987). Isolation and characterization of a cDNA coding for human myeloperoxidase. *Arch. Biochem. Biophys.* 255, 147-155. Hashinaka,K. *et al.* (1988). Multiple species of myeloperoxidase messenger RNAs produced by alternative splicing and differential polyadenylation. *Biochemistry* 27, 5906-5914. Erratum: *Biochemistry* 27, 9226.
1829	M22960 13-JUL-1988	PPGB Protective protein for beta-galactosidase (galactosialidosis)	Galjart,N.J. *et al.* (1988). Expression of cDNA encoding the human 'protective protein' associated with lysosomal beta-galactosidase and neuraminidase: Homology to yeast proteases. *Cell* 54, 755-764.
1834	M23197	CD33 CD33 antigen (differentiation antigen)	Simmons,D. and Seed,B. (1988). Isolation of a cDNA encoding CD33, a differentiation antigen of myeloid progenitor cells. *J. Immunol.* 141, 2797-2800.

Gene Index	Genbank or Affymetrix Accession Number with Submission Date	Gene Description	Reference
1882	M27891 29-SEP-89	CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)	Saitoh,E. *et al.*(1989). The human cystatin C gene (CST3) is a member of the cystatin gene family which is localized on chromosome 20. *Biochem. Biophys. Res. Commun.*162, 1324-1331.
1902	M29474 20-OCT-1989	Recombination activating protein (RAG-1) gene	Schatz,D.G. *et al.*(1989) The V(D)J recombination activating gene, RAG-1. *Cell* 59, 1035-1048.
2111	M62762	ATP6C Vacuolar H+ ATPase proton channel subunit	Gillespie,G.A. *et al.*(1991). CpG island in the region of an autosomal dominant polycystic kidney disease locus defines the 5' end of a gene encoding a putative proton channel. *Proc. Natl. Acad. Sci. U.S.A.* 88, 4289-4293.
2121	M63138	CTSD Cathepsin D (lysosomal aspartyl protease)	Redecker,B. *et al.* (1991). Molecular organization of the human cathepsin D gene. *DNA Cell Biol.* 10, 423-431.
2128	M63379	CLU Clusterin (complement lysis inhibitor; testosterone-repressed prostate message 2; apolipoprotein J)	Wong,P. *et al.* (1993). Genomic organization and expression of the rat TRPM-2 (clusterin) gene, a gene implicated in apoptosis. *J. Biol. Chem.* 268, 5021-5031. Wong,P. *et al.* (1994). Molecular characterization of human TRPM-2/clusterin, a gene associated with sperm maturation, apoptosis and neurodegeneration. *Eur. J. Biochem.* 221, 917-925.
2137	M63835	HIGH AFFINITY IMMUNOGLOBULIN GAMMA FC RECEPTOR I "A FORM" PRECURSOR	van de Winkel,J.G.J. *et al.* (1991). Gene organization of the human high affinity receptor for IgG, Fc-gamma-RI (CD64): Characterization and evidence for a second gene. *J. Biol. Chem.* 266, 13449-13455.
2242	M80254	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE, MITOCHONDRIAL PRECURSOR	Bergsma,D.J. *et al.* (1991). The cyclophilin multigene family of peptidyl-prolyl isomerases. Characterization of three separate human isoforms. *J. Biol. Chem.* 266, 23204-23214.
2288	M84526	DF D component of complement (adipsin)	White,R.T. *et al.* (1992). Human adipsin is identical to complement factor D and is expressed at high levels in adipose tissue. *J. Biol. Chem.* 267, 9210-9213.
2363	M93056	LEUKOCYTE ELASTASE INHIBITOR	Remold-O'Donnell,E. *et al.* (1992). Sequence and molecular characterization of human monocyte/neutrophil elastase inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5635-5639.

Gene Index	Genbank or Affymetrix Accession Number with Submission Date	Gene Description	Reference
2402	M96326	Azurocidin gene	Morgan,J.G. *et al.* (1991). Cloning of the cDNA for the serine protease homolog CAP37/azurocidin, a microbicidal and chemotactic protein from human granulocytes. *J. Immunol.* 147, 3210-3214. Zimmer,M. *et al.* (1992). Three human elastase-like genes co-ordinately expressed in the myelo-monocyte lineage are organized as a single genetic locus on 19pter. *Proc. Natl. Acad. Sci. U.S.A.* 89, 8215-8219.
2546	S82470	BB1=malignant cell expression-enhanced gene/tumor progression-enhanced gene	Fukunaga-Johnson,N. *et al.* (1996). Molecular analysis of a gene, BB1, overexpressed in bladder and breast carcinoma. *Anticancer Res.* 16, 1085-1090.
2565	U00672 10-AUG-1993	IL10R Interleukin 10 receptor	Liu,Y. *et al.* (1994). Expression cloning and characterization of a human IL-10 receptor. *J. Immunol*, 1821-1829.
2800	U14971 21-SEP-1994	RPS9 Ribosomal protein S9	Frigerio,J.M. *et al.* (1995). Cloning, sequencing and expression of the L5, L21, L27a, L28, S5, S9, S10 and S29 human ribosomal protein mRNAs. *Biochim. Biophys. Acta* 1262, 64-68.
3183	U41635 30-NOV-1995	OS-9 precursor mRNA	Su,Y.A. *et al.* (1996). Complete sequence analysis of a gene (OS-9) ubiquitously expressed in human tissues and amplified in sarcomas. *Mol. Carcinog.* 15, 270-275.
3252	U46499 18-JAN-1996	GLUTATHIONE S-TRANSFERASE, MICROSOMAL	DeJong,J.L. *et al.* (1988). Gene expression of rat and human microsomal glutathione S-transferases. *J. Biol. Chem.* 263, 8430-8436. Kelner,M.J. *et al.* (1996). Structural organization of the human microsomal glutathione S-transferase gene (GST12). *Genomics* 36, 100-103.
3258	U46751 19-JAN-1996	Phosphotyrosine independent ligand p62 for the Lck SH2 domain mRNA	Joung,I. *et al.*(1996). Molecular cloning of a phosphotyrosine-independent ligand of the p56lck SH2 domain. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5991-5995.
3320	U50136 27-FEB-1996	Leukotriene C4 synthase (LTC4S) gene	Penrose,J.F. *et al.* (1996). Molecular cloning of the gene for human leukotriene C4 synthase. Organization, nucleotide sequence, and chromosomal localization to 5q35. *J. Biol. Chem.* 271, 11356-11361.
3482	U60319 10-JUN-1996	HLA-H MHC protein HLA-H (hereditary haemochromatosis)	Feder,J.N. *et al.*(1996). A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nature Genet.* 13, 399-408.

Gene Index	Genbank or Affymetrix Accession Number with Submission Date	Gene Description	Reference
3525	U63289 08-JUL-1996	RNA-binding protein CUG-BP/hNab50 (NAB50) mRNA	Timchenko,L.T. *et al.* (1996). Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res.* 24, 4407-4414.
3581	U66580 12-AUG-1996	Putative G protein-coupled receptor (GPR21) gene	O'Dowd,B.F. *et al.* (1997). Cloning and chromosomal mapping of four putative novel human G-protein-coupled receptor genes. *Gene* 187, 75-81.
3820	U81554 10-DEC-1996	CaM kinase II isoform mRNA	Breen,M.A. and Ashcroft,S.J.H. (1997). A truncated isoform of Ca2+/calmodulin-dependent protein kinase II expressed in human islets of Langerhans may result from trans-splicing. *FEBS Lett.* 409, 375-379.
3847	U82759 19-DEC-1996	Homeodomain protein HoxA9 mRNA	Rozenfeld,S. *et al.* Human HOXA9 homeobox cDNA sequence. Unpublished.
4190	X16706 30-OCT-1989	FOS-RELATED ANTIGEN 2	Matsui,M. *et al.* (1990). Isolation of human fos-related genes and their expression during monocyte-macrophage differentiation. *Oncogene* 5, 249-255.
4196	X17042 29-JAN-1990	PRG1 Proteoglycan 1, secretory granule	Stellrecht,C.M. and Saunders,G.F. (1989). Nucleotide sequence of a cDNA encoding a hemopoietic proteoglycan core protein. *Nucleic Acids Res.* 17, 7523.
4229	X52056 07-MAR-1990	SPI1 Spleen focus forming virus (SFFV) proviral integration oncogene spi1	Ray,D. *et al.*(1990). The human homologue of the putative proto-oncogene Spi-1: characterization and expression in tumors. *Oncogene* 5, 663-668.
4322	X59065 16-APR-1991	FGF1 Fibroblast growth factor 1 (acidic) (alternative products)	Wang,W.P. *et al.* (1991). Cloning and sequence analysis of the human acidic fibroblast growth factor gene and its preservation in leukemia patients. *Oncogene* 6, 1521-1529.
4328	X59417 08-MAY-1991	PROTEASOME IOTA CHAIN	Bey,F. *et al.* (1993). The prosomal RNA-binding protein p27K is a member of the alpha-type human prosomal gene family. *Mol. Gen. Genet.* 237, 193-205.
4366	X61587 25-SEP-1991	ARHG Ras homolog gene family, member G (rho G)	Vincent,S. *et al.* (1992). Growth-regulated expression of rhoG, a new member of the ras homolog gene family. *Mol. Cell. Biol.* 12, 3138-3148.
4377	X62654 17-OCT-1991	ME491 gene extracted from H.sapiens gene for Me491/CD63 antigen	Hotta,H. *et al.* (1992). Genomic structure of the ME491/CD63 antigen gene and functional analysis of the 5'-flanking regulatory sequences. *Biochem. Biophys. Res. Commun.* 185, 436-442.

Gene Index	Genbank or Affymetrix Accession Number with Submission Date	Gene Description	Reference
4499	X70297 04-FEB-1993	CHRNA7 Cholinergic receptor, nicotinic, alpha polypeptide 7	Peng,X. *et al.* (1994). Human alpha 7 acetylcholine receptor: cloning of the alpha 7 subunit from the SH-SY5Y cell line and determination of pharmacological properties of native receptors and functional alpha 7 homomers expressed in Xenopus oocytes. *Mol. Pharmacol.* 45, 546-554.
4760	X89066 06-JUL-1995	TRPC1 Transient receptor potential channel 1	Wes,P.D. *et al.* (1995). TRPC1, a human homolog of a Drosophila store-operated channel. *Proc. Natl. Acad. Sci. U.S.A.* 92, 9652-9656.
4847	X95735 16-FEB-1996	Zyxin	Zumbrunn,J. and Trueb,B. (1996). A zyxin-related protein whose synthesis is reduced in virally transformed fibroblasts. *Eur. J. Biochem.* 241, 657-663.
5107	Z29067 13-DEC-1993	Nek3 mRNA for protein kinase	Schultz,S.J. and Nigg,E.A. (1993). Identification of 21 novel human protein kinases, including 3 members of a family related to the cell cycle regulator nimA of Aspergillus nidulans. *Cell Growth Differ.* 4, 821-830. Schultz,S.J. *et al.*(1994). Cell cycle-dependent expression of Nek2, a novel human protein kinase related to the NIMA mitotic regulator of Aspergillus nidulans. *Cell Growth Differ.* 5, 625-635.
5175	Z49269 18-MAY-1995	Chemokine HCC-1	Pardigol,A. *et al.* Nucleotide Sequence of the Gene for the Human Chemokine HCC-1. Unpublished
5280	J02783 15-DEC-1988	P4HB Procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide (protein disulfide isomerase; thyroid hormone binding protein p55)	Cheng,S.Y. *et al.* (1987). The nucleotide sequence of a human cellular thyroid hormone binding protein present in endoplasmic reticulum. *J. Biol. Chem.* 262, 11221-11227.
5318	L43576	(clone EST02946) mRNA May 6 1998	Timms,K.M. *et al.*(1995). 130 kb of DNA sequence reveals two new genes and a regional duplication distal to the human iduronate-2-sulfate sulfatase locus. *Genome Res.* 5, 71-8.

Gene Index	Genbank or Affymetrix Accession Number with Submission Date	Gene Description	Reference
5432	U73936 10-OCT-1996	Soluble protein Jagged mRNA, partial cds	Lindsell,C.E. *et al.* (1995). Jagged: a mammalian ligand that activates Notch1. *Cell* 80, 909-917. Li,L. *et al.* (1997). Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nature Genet.* 16, 243-251. Li,L. *et al.* (1998). The human homolog of rat Jagged1 expressed by marrow stroma inhibits differentiation of 32D cells through interaction with Notch1. *Immunity* 8, 43-55.
5683	U19713 10-JAN-1995	Allograft inflammatory factor-1 (AIF-1) mRNA	Utans,U. *et al.* (1996). Allograft inflammatory factory-1. A cytokine-responsive macrophage molecule expressed in transplanted human hearts. *Transplantation* 61, 1387-1392.
5833	U05572 25-JAN-1994	MANB Mannosidase alpha-B (lysosomal)	Nebes,V.L. and Schmidt,M.C. (1994). Human lysosomal alpha-mannosidase: isolation and nucleotide sequence of the full-length cDNA. *Biochem. Biophys. Res. Commun.* 200, 239-245 Emiliani,C *et al.*(1995). Partial sequence of the purified protein confirms the identity of cDNA coding for human lysosomal alpha-mannosidase B. *Biochem. J.* 305 (Pt 2), 363-366.
5955	U50327 29-FEB-1996	Protein kinase C substrate 80K-H gene (PRKCSH)	Ophoff,R.A. *et al.* A 3 Mb region for the FHM locus on 19p13.1-p13.2; exclusion of PRKCSH as a candidate gene. Unpublished
6005	M32304 23-FEB-1990	TIMP2 Tissue inhibitor of metalloproteinase 2	Boone,T.C. *et al.*(1990). cDNA cloning and expression of a metalloproteinase inhibitor related to tissue inhibitor of metalloproteinases. *Proc. Natl. Acad. Sci. U.S.A.* 87, 2800-2804.
6041	L09209	APLP2 Amyloid beta (A4) precursor-like protein 2	Sprecher,C.A. *et al.* (1993). Molecular Cloning of the cDNA for a Human Amyloid Precursor Protein Homolog (APPH). *Biochemistry* 32, 4481-4486.
6141	Y08765 10-OCT-1996	ZFM1 protein alternatively spliced product	Arning,S. *et al.* (1996). Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA* 2, 794-810.
6185	X64072 05-MAR-1992	SELL Leukocyte adhesion protein beta subunit	Weitzman,J.B. *et al.* (1991). The gene organisation of the human beta 2 integrin subunit (CD18). *FEBS Lett.* 294, 97-103.

Gene Index	Genbank or Affymetrix Accession Number with Submission Date	Gene Description	Reference
6201	Y00787 03-MAY-1988	INTERLEUKIN-8 PRECURSOR	Matsushima,K. *et al.* (1988). M lecular cloning of a human monocyte-derived neutrophil chemotactic factor (MDNCF) and the induction of MDNCF mRNA by interleukin 1 and tumor necrosis factor. *J. Exp. Med.* 167, 1883-1893.
6218	M27783	ELA2 Elastase 2, neutrophil	Farley,D. *et al.* (1988). Molecular cloning of human neutrophil elastase. *Biol. Chem. Hoppe-Seyler* 369, 3-7.
6373	M81695	ITGAX Integrin, alpha X (antigen CD11C (p150), alpha polypeptide)	Corbi,A.L. *et al.* (1987). cDNA cloning and complete primary structure of the alpha subunit of a leukocyte adhesion glycoprotein, p150,95. *EMBO J.* 6, 4023-4028.
6376	M83652	PFC Properdin P factor, complement	Nolan,K.F. *et al.* (1991). Molecular cloning of the cDNA coding for properdin, a positive regulator of the alternative pathway of human complement. *Eur. J. Immunol.* 21, 771-776. Weiler,J.M. and Maves,K.K. (1992). Detection of properdin mRNA in human peripheral blood monocytes and spleen. *J. Lab. Clin. Med.* 120, 762-766.
6378	M83667	NF-IL6-beta protein mRNA	Kinoshita,S. *et al.* (1992). A member of the C/EBP family, NF-IL6 beta, forms a heterodimer and transcriptionally synergizes with NF-IL6. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1473-1476.
6502	U31973 21-JUL-1995	Phosphodiesterase A' subunit (PDE6C) mRNA	Piriev,N.I. *et al.* (1995). Gene structure and amino acid sequence of the human cone photoreceptor cGMP-phosphodiesterase alpha' subunit (PDE6A2) and its chromosomal localization to 10q24. *Genomics* 28, 429-435. Vicztan,A.S. *et al.* (1995). Isolation and characterization of a cDNA encoding the alpha subunit of human cone cGMP-phosphodiesterase. *Gene* 166, 205-211.
6563	U51333 14-MAR-1996	HK3 Hexokinase 3 (white cell)	Furuta,H. *et al.*(1996). Sequence of human hexokinase III cDNA and assignment of the human hexokinase III gene (HK3) to chromosome band 5q35.2 by fluorescence in situ hybridization. *Genomics* 36, 206-209.
6584	Z54367 12-OCT-1995	GB DEF = Plectin	Liu,C.G. *et al.* (1996). Human plectin: organization of the gene, sequence analysis, and chromosome localization (8q24). *Proc. Natl. Acad. Sci. U.S.A.* 93, 4278-4283.

Gene Index	Genbank or Affymetrix Accession Number with Submission Date	Gene Description	Reference
6797	J03801 27-OCT-1988	LYZ Lysozyme	Chung,L.P. *et al.*(1988). Cloning the human lysozyme cDNA: inverted Alu repeat in the mRNA and in situ hybridization for macrophages and Paneth cells. *Proc. Natl. Acad. Sci. U.S.A.* 85, 6227-6231.
6803	M1904	LYZ Lysozyme	Yoshimura,K. *et al.* (1988). Human lysozyme: sequencing of a cDNA, and expression and secretion by Saccharomyces cerevisiae. *Biochem. Biophys. Res. Commun.* 150, 794-801.
6806	X14008 18-JAN-1989	Lysozyme gene (EC 3.2.1.17)	Peters,C.W. *et al.*(1989). The human lysozyme gene. Sequence organization and chromosomal localization. *Eur. J. Biochem.* 182, 507-516.
6919	X16546 18-SEP-1989	RNS2 Ribonuclease 2 (eosinophil-derived neurotoxin; EDN)	Hamann,K.J. *et al.* (1990). Structure and chromosome localization of the human eosinophil-derived neurotoxin and eosinophil cationic protein genes: evidence for intronless coding sequences in the ribonuclease gene superfamily. *Genomics* 7, 535-546.

Appendix C: Associations Predicting AML Treatment Outcome

Treatment Outcome Predictors

Clustered Raw Data

Matthews Relation	Observed	Association
0.8324	5 Successful Treatment	Value:B Gene:400 where A=1014.757576 B=453.384615 Value:B Gene:720 where A=269.704545 B=38.107143 Value:A Gene:1885 where A=-7.880952 B=100.633333 Value:A Gene:2800 where A=14898.151515 B=9699.666667 Value:A Gene:3525 where A=1.277778 B=-167.500000 Value:A Gene:3581 where A=-95.976190 B=-8.400000 Value:A Gene:3820 where A=842.880000 B=158.574468 Value:B Gene:4760 where A=-55.972222 B=34.444444 Value:A Gene:5175 where A=-5.350000 B=-334.093750 Value:A Gene:5318 where A=9.441176 B=132.788474 Value:A Gene:5955 where A=851.200000 B=266.340426
0.8324	5 Successful Treatment	Value:D Gene:1436 where A=101.708 B=670.625 C=439 D=320.214 E=-200.333 F=229.692 G=-28 Value:C Gene:3847 where A=707.2 B=182.091 C=1145.29

What Is Claimed Is:

1. A method for diagnosing acute lymphoblastic leukemia (ALL), comprising:

(a) measuring the levels of gene expression of leukotriene C4 synthase (LTC4S) gene and Zyxin in a biological sample taken from a patient suspected of having ALL; and

(b) comparing the levels of gene expression in said biological sample with a standard sample, wherein low levels of expression are indicative of a diagnosis of ALL.

2. A method for diagnosing ALL, comprising:

(a) measuring the levels of gene expression of LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog, PPGB Protective protein for beta-galactosidase, and Zyxin in a biological sample taken from a patient suspected of having ALL; and

(b) comparing the levels of gene expression in said biological sample with a standard sample, wherein low levels of expression are indicative of a diagnosis of ALL.

3. A method for determining a prognosis of a patient with AML, comprising:

(a) measuring the levels of gene expression of POU3F1 POU domain, class 3, transcription factor 1 and GB DEF = homeodomain protein HoxA9 mRNA in a biological sample taken from a patient with AML; and

(b) comparing the levels of gene expression in said biological sample with a standard sample, wherein medium-high levels of POU3F1 POU domain, class 3, transcription factor 1 and high levels of GB DEF = homeodomain protein HoxA9 mRNA, are indicative of a favorable prognosis.

4. A method for screening drugs which are useful for treating acute leukemia, comprising:

(a) administering to a cell culture a drug of interest;

(b) comparing the levels of gene expression of leukotriene C4 synthase (LTC4S) gene and/or Zyxin before administration of said drug with the levels of gene expression after administration of said drug, wherein a modulation of gene expression level after administration of the drug is indicative of a drug useful for treating acute leukemia.

5. A method for screening drugs which are useful for treating acute leukemia, comprising:

(a) administering to a cell culture a drug of interest; and

(b) comparing the levels of gene expression of LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog, PPGB Protective protein for beta-galactosidase, and/or Zyxin before administration of said drug with the levels of gene expression after administration of said drug, wherein a modulation of gene expression level after administration of the drug is indicative of a drug useful for treating acute leukemia.

6. A kit for diagnosing ALL, comprising:

(a) a means for measuring gene expression of leukotriene C4 synthase (LTC4S) gene; and

(b) a means for measuring gene expression of Zyxin.

7. A kit for diagnosing ALL, comprising:

(a) a means for measuring gene expression of LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog;

(b) a means for measuring gene expression of PPGB Protective protein for beta-galactosidase; and

(c) a means for measuring gene expression of Zyxin.

8. A method for screening drugs which are useful for treating acute leukemia, comprising:

(a) administering to a cell culture a drug of interest; and

(b) comparing the levels of gene expression of POU3F1 POU domain, class 3, transcription factor 1 and/or GB DEF = homeodomain protein HoxA9 mRNA in a biological sample taken from a patient with acute leukemia, wherein a modulation of gene expression level after administration of the drug is indicative of a drug useful for treating acute leukemia.

9. The use of gene expression levels of leukotriene C4 synthase (LTC4S) gene and Zyxin to diagnose ALL.

10. The use of gene expression levels of LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog, PPGB Protective protein for beta-galactosidase, and Zyxin to diagnose ALL.

11. The use of gene expression levels of POU3F1 POU domain, class 3, transcription factor 1 and GB DEF = homeodomain protein HoxA9 mRNA for the prognosis of AML.

12. A method for diagnosing acute myeloid leukemia (AML), comprising:

(a) measuring the levels of gene expression of Zyxin and ELA2 Elastase 2, neutrophil, in a biological sample taken from a patient suspected of having AML; and

(b) comparing the levels of gene expression in said biological sample with a standard sample, wherein high levels of expression are indicative of a diagnosis of AML.

Methods for the Diagnosis and Prognosis of Acute Leukemias

Abstract

The present invention relates to the diagnosis of the distinction between acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) and prognosis of AML. Disclosed is a means to diagnose the distinction between ALL and AML employing measurement of the abundance of the nucleic acid or protein products of small combinations (two, three or more) of particular human genes. The invention further describes the use of the measurement of the abundance of the nucleic acid or protein product of two human genes for prognostic indication in AML. The invention also relates to therapies targeted at these indicator genes, and the screening of drugs for cancer that target these indicator genes or their protein products.

G:\DOCUMENT\CAROL\GENERAL\MISC\A281-311.wpd